STAT 5200 Handout #4: Tests and Distributions

Example: [Beet Lice] A simple experiment is conducted to evaluate the efficacy of four chemical treatments for beet lice. 100 beet plants in individual pots are used in the experiment. 25 plants are randomly assigned to each of the chemical treatments. The plots are spatially separated (in this random order) and the chemical treatments are applied. At the end of two weeks, the number of lice on each plant is counted.

(to each pot/plant separately) — *response variable*

First: random assignment
Second: analysis (for now, just two treatments)
Experimental and measurement units:

→ plants or pots *→ plants or pots*

delete in SAS Studio

```
/* Define options */
ods html image_dpi=300 style=journal;


/* How to randomize? */
/* First create an unrandomized design */
data Unrandomized;
   do ID=1 to 100;
     output;
   end;
data Unrandomized; set Unrandomized;
  Chemical = 'A';
  if ID > 25 then Chemical = 'B';
  if ID > 50 then Chemical = 'C';
  if ID > 75 then Chemical = 'D';
run;
proc print data=Unrandomized;
  title1 'Unrandomized Design';
run;
/* Then randomize the design */
proc plan seed=1234;
   factors ID=100;
   output data=Unrandomized out=Randomized;
run;
/* See the randomized design */
proc sort data=Randomized;
   by ID;
proc print data=Randomized;
  title1 'Randomized Design';
run;
```

makes randomization reproducible (seed=1234)

Unrandomized Design

Obs	ID	Chemical
1	1	A
2	2	A
…		
99	99	D
100	100	D

Randomization
The PLAN Procedure

Factor	Select	Levels	Order
ID	100	100	Random

ID					
25	10	40	13	…	62

Randomized Design

Obs	ID	Chemical
1	1	C
2	2	A
…		
99	99	B
100	100	D

```
/* Enter data */
data lice;
  input Chemical $ NumLice @@;
  label Chemical = "Chemical Treatment";
  label NumLice = "Number of Beet Lice";
  cards;
 A 12   B 10   C 23   D 32   A 13   B 21   C 14   D 26   A 26   B 34
 C 14   D 24   A 13   B 15   C 20   D 16   A 17   B  5   C 27   D 32
 A 24   B 22   C 25   D 18   A 14   B 12   C 17   D 33   A 10   B 25
 C 18   D 16   A  6   B 18   C 29   D 34   A  4   B 12   C 14   D 18
 A  2   B  2   C 31   D  9   A 10   B  2   C  5   D 19   A  8   B 10
 C 13   D 29   A  6   B 22   C 18   D 30   A  7   B 17   C 23   D 18
 A 13   B 20   C 16   D 25   A 18   B 19   C 13   D 20   A 10   B 20
 C 23   D 21   A 18   B 12   C  4   D 27   A  3   B 11   C 16   D 31
 A  4   B 16   C 17   D 25   A 18   B  5   C  9   D 33   A 13   B 11
 C 28   D 24   A 10   B 17   C 23   D 16   A 21   B 16   C 19   D 24
 ;
run;


/* Look at numerical data summaries */
proc means data=lice mean;
 class Chemical;
 var NumLice;
 title1 'Mean comparison';
 title2 '(and num. of exp. units per treatment)';
run;
```

list factors ← (class)

Mean comparison
(and num. of exp. units per treatment)

The MEANS Procedure

Analysis Variable : NumLice
Number of Beet Lice

Chemical Treatment	N Obs	Mean
A	25	12.0000000
B	25	14.9600000
C	25	18.3600000
D	25	24.0000000

$= \bar{y}_A$

$= \bar{y}_B$

```
/* Compare chemicals A & B */
data liceAB; set lice;
  if Chemical = 'A' | Chemical = 'B';
proc ttest data=liceAB;
  class Chemical;
  var Numlice;
  title1 'Two-sample t-test of A vs. B';
run;
```

Two-sample t-test of A vs. B

The TTEST Procedure

Variable: NumLice (Number of Beet Lice)

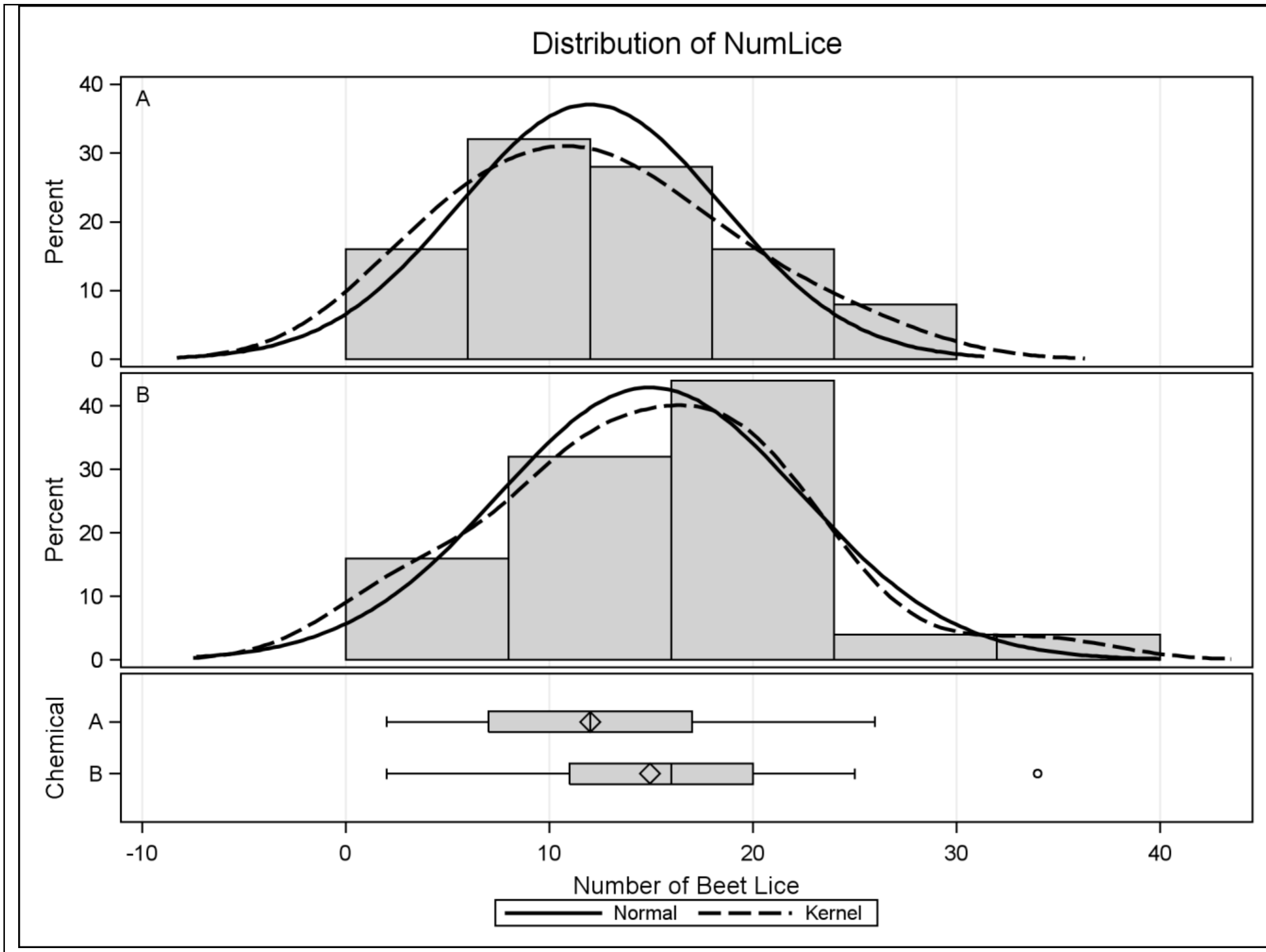
Chemical	N	Mean	Std Dev	Std Err	Minimum	Maximum
A	25	12.0000	6.4550	1.2910	2.0000	26.0000
B	25	14.9600	7.4357	1.4871	2.0000	34.0000
Diff (1-2)		-2.9600	6.9626	1.9693		

Chemical	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
A		12.0000	9.3355	14.6645	6.4550	5.0402	8.9798
B		14.9600	11.8907	18.0293	7.4357	5.8060	10.3442
Diff (1-2)	Pooled	-2.9600	-6.9196	0.9996	6.9626	5.8063	8.6984
Diff (1-2)	Satterthwaite	-2.9600	-6.9216	1.0016			

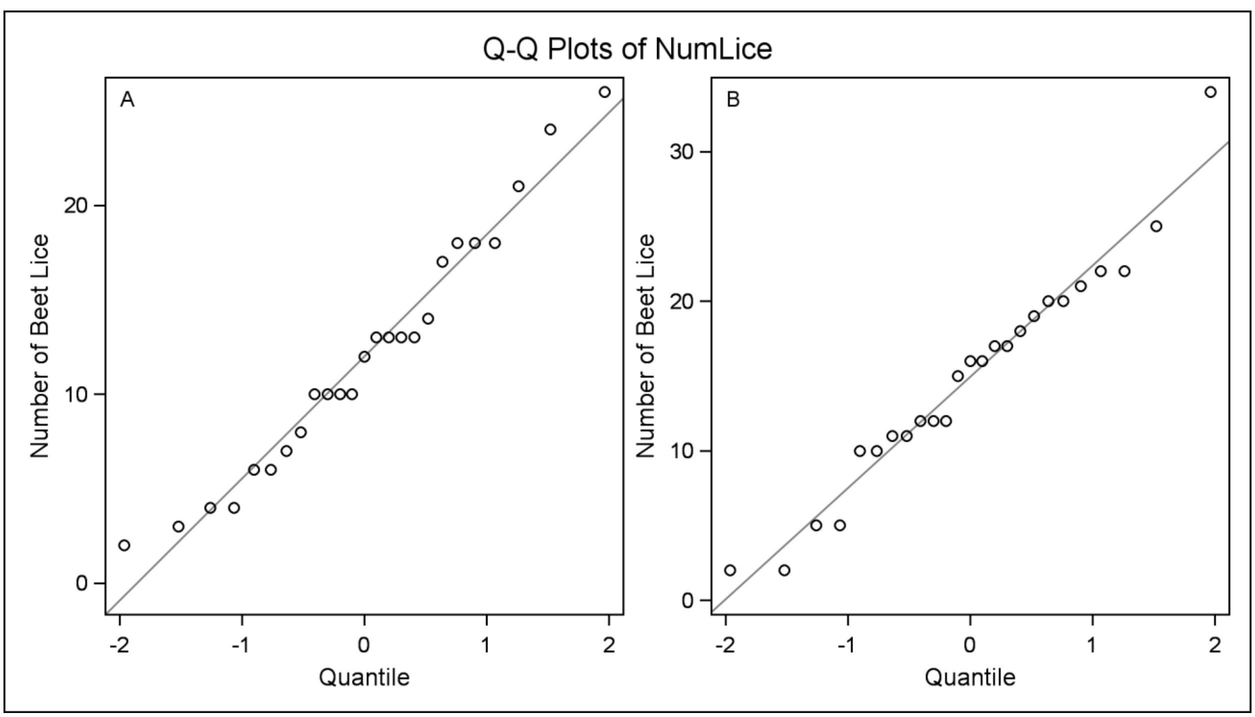
Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	48	-1.50	0.1394
Satterthwaite	Unequal	47.071	-1.50	0.1395

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	24	24	1.33	0.4936



Distribution of NumLice
A
B
Percent
Chemical
Number of Beet Lice
Normal
Kernel



Q-Q Plots of NumLice
A
B
Number of Beet Lice
Quantile

```
/* Numerical check for normality */
proc sort data=liceAB;
 by Chemical;
run;
proc univariate data=liceAB normal;
 by Chemical;
 var NumLice;
 title1 'Check normality';
run;
```

Check normality

The UNIVARIATE Procedure
Variable: NumLice (Number of Beet Lice)
Chemical Treatment=A

Tests for Normality

Test	Statistic		p Value	
Shapiro-Wilk	W	0.963181	Pr < W	0.4814
Kolmogorov-Smirnov	D	0.118442	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.04718	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.304092	Pr > A-Sq	>0.2500

Check normality

The UNIVARIATE Procedure
Variable: NumLice (Number of Beet Lice)
Chemical Treatment=B

Tests for Normality

Test	Statistic		p Value	
Shapiro-Wilk	W	0.967845	Pr < W	0.5910
Kolmogorov-Smirnov	D	0.094714	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.036261	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.27724	Pr > A-Sq	>0.2500

```
/* Look at a different two-sample test */
proc npar1way wilcoxon data=liceAB;
  class Chemical;
  var NumLice;
  title1 'Wilcoxon Rank-Sum Test';
run;
```

```
                              Wilcoxon Rank-Sum Test


                         t Approximation
                         One-Sided Pr <  Z       0.0825
                         Two-Sided Pr > |Z|      0.1650
```

```
/* See hyper-sensitivity of numerical normality tests */
data temp; do i=1 to 40000; output;  end;
data temp; set temp;  x = rannor(1234);
proc univariate data=temp normal;
 histogram x / normal(color=blue) cfill=yellow;
 probplot x / normal(mu=est sigma=est color=red w=3 L=2);
 title1 'Sensitivity of numerical test for normality';
run;
```

Sensitivity of numerical test for normality

Tests for Normality

Test		Statistic		p Value
Kolmogorov-Smirnov	**D**	0.005116	**Pr > D**	0.0122
Cramer-von Mises	**W-Sq**	0.202149	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	1.077539	**Pr > A-Sq**	0.0083